

My second toolbox presentation was on my research in the Northeastern Astrophysics Group with Dr. Bryanne McDonough and recent graduate Sathvika Iyengar. Our work is investigating which astrophysical parameters (e.g. black hole mass) dictate galaxy quenching. A galaxy is considered quenched if its star formation rate is below a certain threshold. We are taking a machine learning approach to the problem, mainly focusing on the random forest classifier in sci-kit learn. The features we are considering include (but are not limited to) the mass of the black hole at the center of the galaxy, the likelihood of galaxy-galaxy interactions (over density), the mass of all the particles in the galaxy's halo (halo mass), and the cumulative energy injected into the galaxy by the black hole. The first step to the research was preprocessing the data so it could be interpreted by the random forest classifier. This involved reading various csv and numpy files into a pandas data frame and filtering out some rows with NaN values and galaxies that are too small. Each row in the DataFrame represents a spatial pixel (spaxel) in the TNG100 cosmological simulation of galaxy evolution. The random forest is built from an ensemble of decision trees and makes a classification based on a majority voting system. Each tree is given a random subset of the data to reduce overfitting. Part of the reason we chose random forest is because of its built in importance scores that show how much each feature contributes to the classification. Initial runs of the random forest indicate that halo mass is most important in low mass galaxies and energy injected into the black hole is most important in high mass galaxies. Unfortunately, these results are illegitimate; some galaxies had some spaxels in the training set and some in the testing set, indicating that the model is data leakage (the testing data is not completely unseen). When we restricted each galaxy to

either the training or testing set, the results worsened as the model overfit. This is the problem we are currently working on fixing.